

# On the Capacity of Interference Channels with One Cooperating Transmitter

Ivana Marić  
Stanford University  
Stanford, CA  
ivanam@wsl.stanford.edu

Andrea Goldsmith  
Stanford University  
Stanford, CA  
andrea@wsl.stanford.edu

Gerhard Kramer  
Bell Labs, Alcatel-Lucent  
Murray Hill, NJ  
gkr@bell-labs.com

Shlomo Shamai (Shitz)  
Technion  
Haifa, Israel  
sshlo@ee.technion.ac.il

**Abstract**—Inner and outer bounds are established on the capacity region of two-sender, two-receiver interference channels where one transmitter knows both messages. The transmitter with extra knowledge is referred to as being cognitive. The inner bound is based on strategies that generalize prior work, and include rate-splitting, Gel'fand-Pinsker coding and cooperative transmission. A general outer bound is based on the Nair-El Gamal outer bound for broadcast channels. A simpler bound is presented for the case in which one of the decoders can decode both messages. The bounds are evaluated and compared for Gaussian channels.

## I. INTRODUCTION AND RELATED WORK

Two-sender, two-receiver channel models allow for various forms of transmitter cooperation. When senders are unaware of each other's messages, we have the interference channel [1], [2]. In wireless networks, the broadcast nature of the wireless medium allows nodes to overhear transmissions and possibly decode parts of other users' messages. An encoder that has such knowledge can use it to improve its own rate and the other user's rate. The level of cooperation and performance improvement will depend on the amount of information the encoders share. In the interference channel, rate gains from the transmitter cooperation were demonstrated in [3].

Channel models with cooperating nodes are of interest also for networks with cognitive users. Cognitive radio [4] technology is aimed at developing smart radios that are both aware of and adaptive to the environment. Such radios can efficiently sense the spectrum, decode information from detected signals and use that knowledge to improve the system performance. This technology motivates information-theoretic models that try to capture the cognitive radio characteristics. In that vein, this paper considers a two-sender, two-receiver channel model in which, somewhat idealistically, we assume that cognitive capabilities allow one user to know the full message of the other encoder, as shown in Fig. 1. Existing encoding schemes can bring different rate gains that depend on the channel characteristics and topology, making it challenging

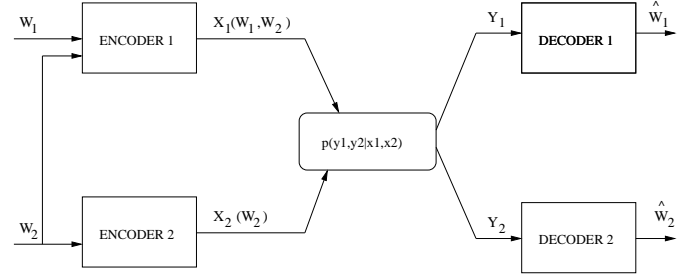


Fig. 1. Interference channel with cooperating encoder.

to determine the capacity even for special cases. This paper is a step along this path. It would further be interesting to extend the existing results to large networks with cooperating encoders.

Investigating the capacity region of the channel model we consider is the focus of much recent work. In particular, the interference channel with one cooperating encoder was dubbed the *cognitive radio channel* and achievable rates were presented in [3], [5]. A general encoding scheme was also proposed more recently in [6]. The capacity region for the Gaussian case of weak interference was determined in [7] and [8]. The results of [7], [8] were extended to the Gaussian MIMO cognitive radio network and shown to achieve the sum-capacity in [9]. Related work can also be found in [10], [11]. However, the conclusions of [9] do not immediately apply to the single-antenna cognitive radio channel. In this paper, we present a scheme that generalizes those in [7]–[12]. The scheme is similar to the one in [6]: as in [6] and [3], an encoder uses *rate-splitting* [2] to enable the other receiver to decode part of the interference; the cognitive transmitter cooperates in sending the other user's message to its intended receivers and uses Gel'fand-Pinsker (GP) binning [13] to reduce interference to its own receiver. The key difference of our contribution to the prior work is in the way the binning is performed. An overview of the encoding scheme is given in the next section. The encoding scheme is derived in Section IV, compared to other results and adapted for Gaussian channel in Section VI.

In Section V, we present two outer bounds for the interference channel with one cooperating encoder. The first bound is based on the Nair-El Gamal broadcast outer bound, [14]. It has the same mutual information expression as the one

<sup>1</sup>The work by I. Marić and A. Goldsmith was supported in part from the DARPA ITMANET program under grant 1105741-1-TFIND, Stanford's Clean Slate Design for the Internet Program and the ARO under MURI award W911NF-05-1-0246. The work of G. Kramer was partially supported by the Board of Trustees of the University of Illinois Subaward No. 04-217 under NSF Grant No. CCR-0325673. The work of S. Shamai was supported by the EU 7th framework program: NEWCOM++.

in [14], the only difference is in the input distribution over which the optimization is performed. The bound thus reflects the resemblance of the considered channel to the broadcast channel (BC), and the difference given by the fact that encoder 2 has only partial knowledge of messages sent in the channel. We then present an outer bound for the strong interference case that is of the same form as the one in [15, Sect.V], and compare it to the achievable rate region in Gaussian channels. Results also demonstrate an improvement compared to the general scheme of [6].

## II. OVERVIEW OF THE ENCODING STRATEGY

The considered channel model has elements of both the interference channel (IC) and the broadcast channel (BC). Encoding techniques developed for either of them are therefore potentially useful. If the message  $W_2$  of encoder 2 was not known at the cognitive encoder, the considered channel would reduce to the interference channel (IC). The best achievable rate region for the IC, [16], is achieved by rate-splitting [2]: each encoder divides its message into two parts and encodes each of them with a separate codebook. This allows receivers to decode one of the two sub-messages of the other user's and cancel a part of the interference that it would otherwise create. Rate-splitting in the cognitive radio channel model was applied in [3], [6]. In this paper, rate-splitting is performed at the cognitive encoder.

Additional knowledge allows the cognitive encoder to employ a number of techniques in addition to rate-splitting. In particular, to improve the rate for the noncognitive communicating pair, the cognitive encoder can *cooperate* by encoding  $W_2$  to help convey it to the other decoder. On the other hand, any signal carrying information about  $W_2$  creates interference to the cognitive encoder's receiver. This interference is known at the cognitive transmitter and the precoding technique, i.e. Gel'fand-Pinsker binning [13] and, specifically, dirty-paper coding (DPC) [17] in Gaussian channels, can be employed. In fact, GP binning is crucial for the cognitive radio channel: together with cooperation, it leads to capacity in certain scenarios, [7], [8], [9]. It is not surprising that DPC brings gains in the Gaussian cognitive radio channel: if the non-cognitive encoder is silent, we have the broadcast channel from the cognitive encoder to two receivers, for which dirty-paper coding is the optimal strategy [18], [19].

In general, however, there are two differences at the cognitive encoder from the classical GP setting. First, the interference carries useful information for receiver 2. Second, the interference is a *codebook* of some rate and can thus have lower entropy than in the GP setting. As we will see in Sect. II-B, the latter can be exploited to achieve a higher rate.

We note that due to rate-splitting, there is a common part of  $W_1$  decoded at the both receivers and precoded against interference. Since the signal carrying this common message experiences *different* interference at the two receivers, we use the ideas of [20] and [21] that respectively extend [13] and [17] to channels with different states non-causally known to

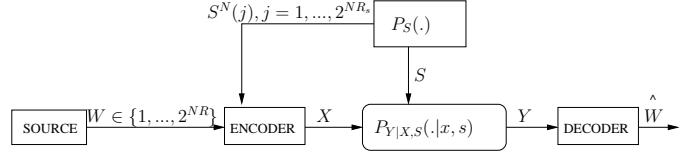


Fig. 2. Communication from the cognitive transmitter to the corresponding receiver.

the encoder. In the Gaussian channel, dirty paper coding is generalized to *carbon-copying* onto dirty paper [21] to adjust to the interference experienced at both receivers.

### A. Summary of Techniques and Special Cases

Although the interference channel with one cooperating encoder can easily be visualized as an extension of the classical IC, a number of techniques become potentially relevant due to additional knowledge of the cognitive encoder:

- Rate splitting at encoder 1: Improves rate  $R_2$  through interference cancellation at decoder 2.
- GP binning and binning against a codebook: Improves rate  $R_1$  by precoding against interference. It also allows decoder 1 to decode message  $W_2$  (or part of it) when  $R_2$  is small, as will be shown in Sect. II-B.
- Carbon-copying onto dirty paper: further improves the rate of the common message sent at the cognitive encoder
- Cooperation: Encoder 1 contributes to rate  $R_2$  by encoding  $W_2$ .

A general encoding scheme that brings these techniques together is described in Section IV. There will be number of special cases for which a subset of techniques will suffice:

- 1) Strong interference: Both decoders can decode both messages with no rate penalty, so there is no need for either rate-splitting or binning. Superposition coding achieves capacity, [15].
- 2) Cognitive encoder decodes both messages: Again, there is no need for binning. Rate-splitting and superposition coding achieve capacity, [22], [23].
- 3) Weak interference at receiver 2: There is no need for common part of message  $W_1$  and hence for rate-splitting. Dirty paper coding and cooperation achieve capacity in Gaussian channel, [7], [8], [9].

### B. Rate Improvement due to Binning Against Codebook

For the communication between the cognitive transmitter and its corresponding receiver, a codebook carrying  $W_2$  creates interference. The situation is depicted in Fig 2, where  $S$  plays the role of the codebook of rate  $R_s$  interfering with the communication of message  $W$  at rate  $R$ . While in the GP problem the interference  $S$  is generated by a discrete memoryless source (DMS), the interference in the cognitive setting is a *codebook* of some rate,  $R_s$ . The next lemma reflects the fact that when  $R_s$  is small, this can be exploited for potential rate gains.

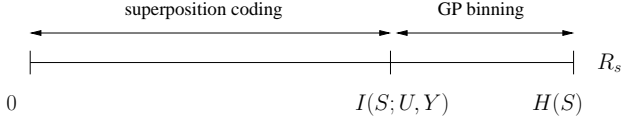


Fig. 3. Binning against a codebook.

**Lemma 1:** For the communication situation of Fig. 3, the rate

$$R \leq \max_{P_{U|S}, f(\cdot)} \min \{I(X; Y|S), \max \{I(U; S; Y) - R_s, I(U; Y) - I(U; S)\}\} \quad (1)$$

is achievable.

For  $I(S; U, Y) \leq R_s \leq H(S)$ , binning achieves the GP rate given by the second term in (1).

For  $R_s \leq I(S; U, Y)$ , superposition coding achieves the rate given by the first term in (1).

The two cases are shown in Fig. 3.

*Proof:* See Appendix B. ■

**Remark 1:** Rate (1) can be written as

$$R \leq \max_{P_{U|S}, f(\cdot)} \{I(X, S; Y) - \max \{I(S; Y) \min \{R_s, I(U, Y; S)\}\}\}. \quad (2)$$

From (1) and (2), we observe that  $I(S; U, Y) \leq R_s \leq H(S)$ , corresponds to the classical GP setting. Potential rate improvement comes for  $R_s \leq I(S; U, Y)$ . Interestingly, in this case the receiver decodes both indexes  $(w, j)$ , thus learning both its message and the interference. A related setting in which both data and the channel state information is communicated to the receiver was analyzed in [24], [25].

In the cognitive setting of Fig. 1, index  $j$  carries information about  $W_2$ . The implication is that, when  $R_s$  is small, receiver 1 will decode a part (or the whole) of  $W_2$  without having encoder 2 rate split to send common information in the sense of [2], [16].

Recall that, due to rate-splitting, encoder 1 uses two codebooks to send a common and a private index. We denote these respective codebooks as  $(U_{1c}^N, U_{1a}^N)$ . We can distinguish four cases depending on whether the two codebooks are generated through binning or superposition coding:

- 1) Binning: Both  $(U_{1c}^N, U_{1a}^N)$  are binned against the codebook of the non-cognitive encoder,  $X_2^N$ .
- 2) Superposition coding: Codebooks are superimposed on  $X_2^N$ .
- 3) Binning then superposition coding:  $U_{1c}^N$  is binned against  $X_2^N$ , and  $U_{1a}^N$  is superimposed on  $(X_2^N, U_{1c}^N)$ .
- 4) Superposition coding then binning:  $U_{1c}^N$  is superimposed on  $X_2^N$ ;  $U_{1a}^N$  is superimposed on  $U_{1c}^N$  and binned against  $X_2^N$ .

In the last two cases, decoder 1 can decode  $W_2$  due to superposition coding of  $U_{1a}^N$  or  $U_{1c}^N$  on  $X_2^N$ , as shown in

Lemma 1. The setting thus corresponds to the *cognitive radio with the degraded message set*. For this channel model, superposition coding achieves the capacity [22], [23]. The two last cases can therefore bring no improvement. The achievable rate region is the union of two rate regions, achieved by binning or superposition coding. We will derive these regions after formally defining the problem in the next section. We remark that in the above encoding scheme, codebook  $U_{1a}^N$  is always superimposed on  $U_{1c}^N$ . The other encoding choice would be to use binning for  $U_{1a}^N$  against the codebook carrying the common message,  $U_{1c}^N$ .

As the final point about the proposed scheme we note that encoder 2 also uses rate-splitting and forms two codebooks  $(X_{2a}^N, X_{2b}^N)$  using superposition coding. Encoder 1 is binning against both codebooks and is not decoding a part of  $W_2$ . An interesting next step would therefore be to choose respective rates  $R_{2a}$  and  $R_{2b}$  following Lemma 1 such that  $(U_{1a}^N, U_{1c}^N)$  are binned against one of the two codebooks, but superimposed on the other. That would facilitate decoding a part of  $W_2$  at receiver 1.

### III. CHANNEL MODEL

Consider a channel with finite input alphabets  $\mathcal{X}_1, \mathcal{X}_2$ , finite output alphabets  $\mathcal{Y}_1, \mathcal{Y}_2$ , and a conditional probability distribution  $p(y_1, y_2 | x_1, x_2)$ , where  $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$  are channel inputs and  $(y_1, y_2) \in \mathcal{Y}_1 \times \mathcal{Y}_2$  are channel outputs. Each encoder  $t$ ,  $t = 1, 2$ , wishes to send a message  $W_t \in \{1, \dots, M_t\}$  to decoder  $t$  in  $N$  channel uses. Message  $W_2$  is also known at encoder 1 (see Fig. 1). The channel is memoryless and time-invariant in the sense that

$$\begin{aligned} p(y_{1,n}, y_{2,n} | x_1^n, x_2^n, y_1^{n-1}, y_2^{n-1}, \bar{w}) \\ = p_{Y_1, Y_2 | X_1, X_2}(y_{1,n}, y_{2,n} | x_{1,n}, x_{2,n}) \end{aligned} \quad (3)$$

for all  $n$ , where  $X_1, X_2$  and  $Y_1, Y_2$  are random variables representing the respective inputs and outputs,  $\bar{w} = [w_1, w_2]$  denotes the messages to be sent, and  $x_t^n = [x_{t,1}, \dots, x_{t,n}]$ . We will follow the convention of dropping subscripts of probability distributions if the arguments of the distributions are lower case versions of the corresponding random variables.

An  $(M_1, M_2, N, P_e)$  code has two encoding functions

$$X_1^N = f_1(W_1, W_2) \quad (4)$$

$$X_2^N = f_2(W_2) \quad (5)$$

two decoding functions

$$\hat{W}_t = g_t(Y_t^N) \quad t = 1, 2 \quad (6)$$

and an error probability

$$P_e = \max\{P_{e,1}, P_{e,2}\} \quad (7)$$

where, for  $t = 1, 2$ , we have

$$P_{e,t} = \sum_{(w_1, w_2)} \frac{1}{M_1 M_2} P[g_t(Y_t^N) \neq w_t | (w_1, w_2) \text{ sent}] \quad (8)$$

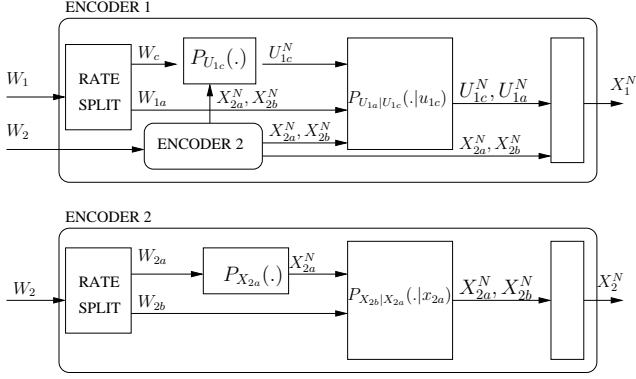


Fig. 4. Encoding structure.

A rate pair  $(R_1, R_2)$  is achievable if, for any  $\epsilon > 0$ , there is an  $(M_1, M_2, N, P_e)$  code such that

$$M_t \geq 2^{NR_t}, \quad t = 1, 2, \text{ and } P_e \leq \epsilon.$$

The capacity region of the interference channel with a cooperating encoder is the closure of the set of all achievable rate pairs  $(R_1, R_2)$ .

#### IV. ACHIEVABLE RATE REGION

To obtain an inner bound, we employ rate splitting. We let

$$R_1 = R_{1a} + R_c \quad (9)$$

$$R_2 = R_{2a} + R_{2b} \quad (10)$$

for nonnegative  $R_{1a}, R_c, R_{2a}, R_{2b}$  which we now specify.

In the encoding scheme, encoder 2 uses superposition coding with two codebooks  $X_{2a}^N, X_{2b}^N$ . Encoder 1 repeats the steps of encoder 2 and adds binning: it encodes the split message  $W_1$  with two codebooks which are Gel'fand-Pinsker precoded against  $X_{2a}^N, X_{2b}^N$ . In particular:

- 1) Binning against  $X_{2a}^N, X_{2b}^N$  is used to create a codebook  $U_{1c}^N$  of common rate  $R_c$ .
- 2) Binning against  $X_{2a}^N, X_{2b}^N$  conditioned on  $U_{1c}$  is used to create a codebook  $U_{1a}^N$  with private rate  $R_{1a}$ .

The encoding structure is shown in Fig. 4.

We have the following result.

*Theorem 1: (joint decoding)* Rates (9)-(10) are achievable if

$$R_{1a} \leq I(U_{1a}; Y_1 | U_{1c}, Q) - I(U_{1a}; X_{2a}, X_{2b} | U_{1c}, Q) \quad (11)$$

$$R_1 \leq I(U_{1c}, U_{1a}; Y_1 | Q) - I(U_{1c}, U_{1a}; X_{2a}, X_{2b} | Q) \quad (12)$$

$$R_2 \leq I(X_2; Y_2, U_{1c} | Q) \quad (13)$$

$$R_2 + R_c \leq I(X_2, U_{1c}; Y_2 | Q) \quad (14)$$

$$R_{2b} \leq I(X_{2b}; Y_2, U_{1c} | X_{2a}, Q) \quad (15)$$

$$R_{2b} + R_c \leq I(X_{2b}, U_{1c}; Y_2 | X_{2a}, Q) \quad (16)$$

for some joint distribution that factors as

$$p(q)p(x_{2a}, x_{2b}, u_{1c}, u_{1a}, x_1, x_2 | q)p(y_1, y_2 | x_1, x_2) \quad (17)$$

and for which all the right-hand sides are nonnegative.

*Proof:* See Appendix A.  $\blacksquare$

*Theorem 2: (sequential decoding)* Rates (9)-(10) are achievable if

$$R_{1a} \leq I(U_{1a}; Y_1 | U_{1c}, Q) - I(U_{1a}; X_2 | U_{1c}, Q) \quad (18)$$

$$R_c \leq \min\{I(U_{1c}; Y_1 | Q), I(U_{1c}; Y_2, X_{2a} | Q)\} - I(U_{1c}; X_2 | Q) \quad (19)$$

$$R_{2a} \leq I(X_{2a}; Y_2 | Q) \quad (20)$$

$$R_{2b} \leq I(X_{2b}; Y_2, U_{1c} | X_{2a}, Q) \quad (21)$$

for some joint distribution that factors as  $p(q)p(x_{2a}, x_{2b}, u_{1c}, u_{1a}, x_1, x_2 | q)p(y_1, y_2 | x_1, x_2)$  and for which the right-hand side of (18) and (19) are nonnegative.  $Q$  is a time sharing random variable.

*Proof:* The proof follows similar steps as the Thm. 1 proof and is omitted. Details can be found in [12].  $\blacksquare$

*Remark 2:* The rates of Thm. 1 include the rates of Thm. 2.

*Remark 3:* Thm. 1 includes the rates of the following schemes:

- The scheme of [7, Thm 3.1] for  $X_{2a} = \emptyset, U_{1c} = \emptyset, X_{2b} = (X_2, U)$  and  $U_{1a} = V$  achieving:

$$R_2 \leq I(X_2, U; Y_2) \quad (22)$$

$$R_1 \leq I(V; Y_1) - I(V; X_2, U) \quad (23)$$

for  $p(u, x_2)p(v | u, x_2)p(x_1 | v)$ .

- The scheme of [26, Lemma 4.2] for  $X_{2a} = \emptyset, X_{2b} = X_2, U_{1a} = \emptyset$ , and  $R_1 = R_c, R_2 = R_{2b}$  as:

$$R_2 \leq I(X_2; Y_2 | U_{1c})$$

$$R_1 \leq \min\{I(U_{1c}; Y_1), I(U_{1c}; Y_2)\}$$

for  $p(x_2)p(u_{1c})$ . The strategy in [26] considers the case when  $I(U_{1c}; Y_1) \leq I(U_{1c}; Y_2)$ .

- Carbon-copy on dirty paper [21] for  $X_{2a} = \emptyset, U_{1a} = \emptyset$ .
- For  $X_{2a} = \emptyset$ , our scheme closely resembles the scheme in [6]. The first difference in our scheme is that two binning steps are not done independently which brings potential improvements. The second difference is in the evaluation of error events.

It is also interesting to compare our scheme to the encoding scheme in [3]. The latter combines rate splitting at both users, with two-step binning at the cognitive user. Each user sends a private index decoded by its receiver, and a common index decoded by both. Again, one difference in our scheme is that two binning steps are not independent. The other is that in our scheme the cognitive encoder cooperates by encoding index  $W_2$ .

The next rate region is obtained by exploiting Lemma 1.

#### A. An Achievable Rate Region with Superposition Coding

Consider a joint distribution (17) and rate  $R_2$  that satisfy

$$R_2 \leq I(X_2; U_{1c}, Y_1) \quad (24)$$

$$R_2 \leq I(X_2; U_{1a}, Y_1 | U_{1c}). \quad (25)$$

From Lemma 1, we know that under respective conditions (24) and (25), superposition of  $U_{1c}^N$  and  $U_{1a}^N$  with  $X_2^N$  should be used instead of binning. The encoding scheme of the cognitive encoder reduces to rate-splitting and superposition coding. The scheme and the obtained rates reduce to that of [6, Thm.5] derived for the *cognitive radio with the degraded message set*, in which the cognitive decoder needs to decode both messages. No rate-splitting at encoder 2 is needed. We restate the result for completeness.

Achievable rates  $(R_1, R_2)$  satisfy

$$\begin{aligned} R_{1a} &\leq I(X_1; Y_1 | X_2, U_{1c}) \\ R_1 &\leq I(X_1; Y_1 | X_2) \\ R_1 + R_2 &\leq I(X_1, X_2; Y_1) \\ R_c + R_2 &\leq I(U_{1c}, X_2; Y_2) \end{aligned} \quad (26)$$

for some joint input distribution  $p(x_2, u_{1c}, x_1)$ .

After Fourier-Motzkin elimination [27], the rates (26) reduce to the following region.

*Theorem 3:* [22]. Achievable rates  $(R_1, R_2)$  satisfy

$$\begin{aligned} R_1 &\leq I(X_1; Y_1 | X_2) \\ R_2 &\leq I(U_{1c}, X_2; Y_2) \\ R_1 + R_2 &\leq I(X_1; Y_1 | X_2, U_{1c}) + I(U_{1c}, X_2; Y_2) \\ R_1 + R_2 &\leq I(X_1, X_2; Y_1) \end{aligned} \quad (27)$$

for some joint input distribution  $p(x_2, u_{1c}, x_1)$ .

*Remark 4:* The above region is the capacity region for the cognitive radio with degraded message sets: the converse follows from [23] where a more general case of confidential messages is analyzed. The result follows by considering the special case of no security.

We have so far presented achievable rates for the cognitive radio channel. We next derive two outer bounds to performance of any encoding scheme in this channel.

## V. OUTER BOUNDS

*Theorem 4:* The set of rate pairs  $(R_1, R_2)$  satisfying

$$R_1 \leq I(V, U_1; Y_1) \quad (28)$$

$$R_2 \leq I(V, U_2; Y_2) \quad (29)$$

$$R_1 + R_2 \leq \min\{I(V, U_1; Y_1) + I(U_2; Y_2 | U_1, V), \quad (30)$$

$$I(U_1; Y_1 | U_2, V) + I(V, U_2; Y_2)\} \quad (31)$$

for input distributions  $p(v, u_1, u_2, x_1, x_2)$  that factor as

$$p(u_1)p(u_2)p(v|u_1, u_2)p(x_2|u_2)p(x_1|u_1, u_2) \quad (32)$$

is an outer bound to the capacity region.

*Proof:* See Appendix C. ■

*Remark 5:* We observe that (28)-(31) is of the same form as the outer bound for the broadcast channel in [14, Sect. 3]. The difference is the factorization of the input distribution.

*Remark 6:* One can restrict attention to distributions (32) where  $X_2$  is a function of  $U_2$  and  $X_1$  is a function of  $(U_1, U_2)$ .

The bounds (28)-(31) can then be written as

$$R_1 \leq I(V, U_1; Y_1) \quad (33)$$

$$R_2 \leq I(V, U_2, X_2; Y_2) \quad (34)$$

$$R_1 + R_2 \leq \min\{I(V, U_1; Y_1) + I(X_1, X_2; Y_2 | U_1, V), \quad (35)$$

$$I(X_1; Y_1 | X_2, U_2, V) + I(V, U_2, X_2; Y_2)\} \quad (36)$$

From (34) and (36), we obtain the outer bound of [7, Thm. 3.2]:

$$R_2 \leq I(X_1; Y_1 | X_2) \quad (37)$$

$$R_2 \leq I(U, X_2; Y_2) \quad (38)$$

$$R_1 + R_2 \leq I(X_1; Y_1 | X_2, U) + I(U, X_2; Y_2) \quad (39)$$

where we used notation  $U = [U_2, V]$  and also added (37) as it follows by standard methods. The probability distribution factors as

$$p(u, x_1, x_2)p(y_1, y_2 | x_1, x_2). \quad (40)$$

Interestingly, (37)-(40) was shown to be tight under weak interference [7, Def. 2.3] and in particular for Gaussian channels with weak interference [7], [8].

The following theorem gives a simple upper bound in strong interference.

*Theorem 5:* For an interference channel with one cooperating encoder satisfying

$$I(X_1; Y_1 | X_2) \leq I(X_1; Y_2 | X_2) \quad (41)$$

for all input distribution  $p(x_1, x_2)$ , the set of rate pairs  $(R_1, R_2)$  satisfying

$$R_1 \leq I(X_1; Y_1 | X_2) \quad (42)$$

$$R_1 + R_2 \leq I(X_1, X_2; Y_2) \quad (43)$$

for all input distributions  $p(x_1, x_2)$  is an outer bound to the capacity region.

*Proof:* See Appendix D. ■

*Remark 7:* The bound (43) reflects the fact that, because decoder 2 experiences strong interference, as given by (41), it can decode  $W_1$  with no rate penalty.

We next compare the outer bound of Thm. 5 to the achievable rates for Gaussian channels.

## VI. GAUSSIAN CHANNEL

To illustrate obtained results more concretely, we next consider the Gaussian interference channel described by

$$Y_1 = X_1 + aX_2 + Z_1 \quad (44)$$

$$Y_2 = bX_1 + X_2 + Z_2 \quad (45)$$

where  $Z_t \sim [0, 1]$  and  $E[X_t^2] \leq P_t$ ,  $t = 1, 2$ . In the case of weak interference, i.e.,  $b \leq 1$ , the capacity region was determined in [7], [8].

We next evaluate the rates of Thm. 1 for the special case  $X_{2a} = \emptyset$  and  $Q = \emptyset$ . Rates of Thm.1 for this case reduce to

$$\begin{aligned} R_{1a} &\leq I(U_{1a}; Y_1 | U_{1c}) - I(U_{1a}; X_2 | U_{1c}) \\ R_1 &\leq I(U_{1c}, U_{1a}; Y_1) - I(U_{1c}, U_{1a}; X_2) \\ R_2 &\leq I(X_2; Y_2, U_{1c}) \\ R_2 + R_c &\leq I(X_2, U_{1c}; Y_2). \end{aligned} \quad (46)$$

To simplify (46), we express the conditional entropies in terms of joint entropies, recall that  $R_1 = R_c + R_{1a}$ , and apply Fourier-Motzkin elimination to obtain

$$\begin{aligned} R_1 &\leq I(U_{1c}, U_{1a}; Y_1) - I(U_{1c}, U_{1a}; X_2) \\ R_2 &\leq I(X_2; Y_2, U_{1c}) \\ R_2 &\leq I(X_2, U_{1c}; Y_2) \\ R_1 + R_2 &\leq I(X_2, U_{1c}; Y_2) + I(U_{1a}; Y_1, U_{1c}) \\ &\quad - I(U_{1a}; X_2, U_{1c}). \end{aligned} \quad (47)$$

It is interesting to evaluate the rates of Thm. 2 achieved with sequential decoding for  $X_{2a} = \emptyset, Q = \emptyset$  as was done for joint decoding in (46). This evaluation results in

$$\begin{aligned} R_{1a} &\leq I(U_{1a}; Y_1 | U_{1c}) - I(U_{1a}; X_2 | U_{1c}) \\ R_c &\leq \min\{I(U_{1c}; Y_1), I(U_{1c}; Y_2)\} - I(U_{1c}; X_2) \\ R_2 &\leq I(X_2; Y_2, U_{1c}). \end{aligned} \quad (48)$$

*Remark 8:* When  $I(U_{1c}; Y_1) \leq I(U_{1c}; Y_2)$ , decoder 2 can decode  $W_1$ . Thus, there is no need to rate split at encoder 1 and we choose  $U_{1a} = \emptyset$ . It follows from (47) and (48) that for this case the same rates can be achieved by sequential decoding or by joint decoding.

*Remark 9:* We observe from (48) that  $R_c$ , being a common rate, is bounded by the worst channel, as reflected by the  $\min\{I(U_{1c}; Y_1), I(U_{1c}; Y_2)\}$  term. If  $I(U_{1c}; Y_1) > I(U_{1c}; Y_2)$ , transmitting  $X_{2a}$  will allow decoder 2 to decode part of  $W_2$  before decoding  $W_c$ . It will also serve as an observation when decoding  $W_c$  as suggested by the expression  $I(U_{1c}; Y_2, X_{2a})$  in (19). This will improve the common rate  $R_c$ .

We evaluated region (47) for

$$X_2 \sim \mathcal{N}[0, P_2], \quad X_{1c} \sim \mathcal{N}[0, \alpha\beta P_1], \quad X_{1a} \sim \mathcal{N}[0, \alpha\bar{\beta}P_1]$$

$$\begin{aligned} U_{1c} &= X_{1c} + \lambda_1 X_2 \\ U_{1a} &= X_{1a} + \lambda_2 X_2 \\ X_1 &= X_{1c} + X_{1a} + \sqrt{\frac{\alpha P_1}{P_2}} X_2 \end{aligned} \quad (49)$$

where  $\mathcal{N}[0, \sigma^2]$  denotes the normal distribution with variance  $\sigma^2$ , and  $0 \leq \alpha, \beta \leq 1$  and  $0 \leq \lambda_1, \lambda_2$ . Parameters  $\alpha$  and  $\beta$  determine the amount of power that the cognitive user dedicates respectively for cooperation  $(1 - \alpha)$  and for sending the common message.

We compared the achievable region (47) to the outer bound of Thm. 5 which in Gaussian channels is given by the following corollary:

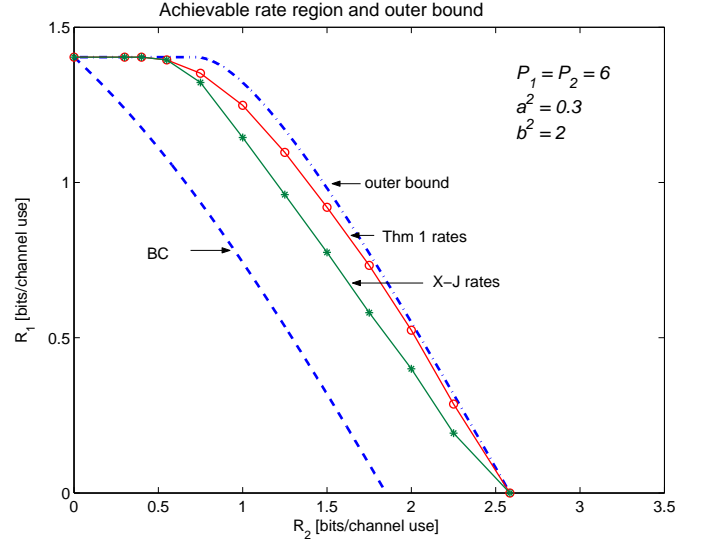


Fig. 5. Achievable rates of Thm. 1 and [6, Thm.5] and outer bound of Cor. 1. Also shown is the capacity region of a BC from the cooperative encoder, i.e. case  $P_2 = 0$ .

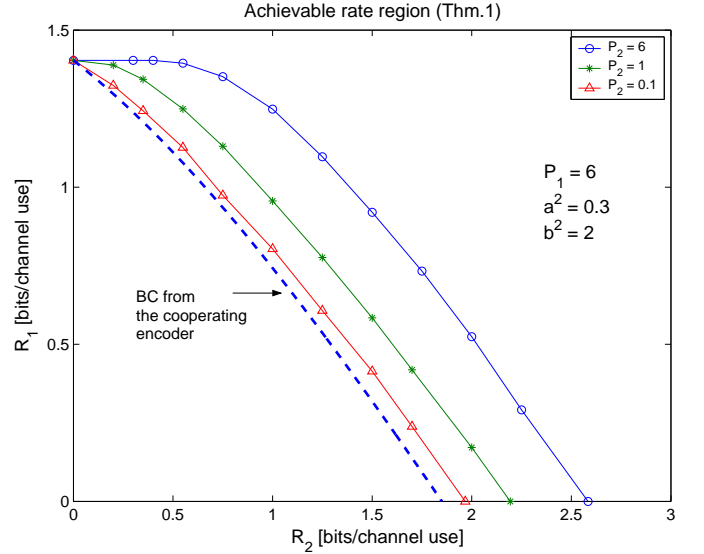


Fig. 6. Achievable rates for different values of the non-cooperating encoder power,  $P_2$ .

*Corollary 1:* When  $b \geq 1$ , any achievable rate pair  $(R_1, R_2)$  satisfies

$$\begin{aligned} R_1 &\leq C((1 - \rho^2)P_1) \\ R_1 + R_2 &\leq C(b^2 P_1 + P_2 + 2\rho\sqrt{b^2 P_1 P_2}) \end{aligned} \quad (50)$$

for some  $\rho$ ,  $0 \leq \rho \leq 1$ , where

$$C(x) = \frac{1}{2} \log(1 + x). \quad (51)$$

Fig. 5 shows the achievable rate region (47) and the outer bound (50) for channel gain values  $a^2 = 0.3, b^2 = 2$  and equal powers  $P_1 = P_2 = 6$ . We observe higher rates of Thm. 1 compared to that of [6, Thm. 5].

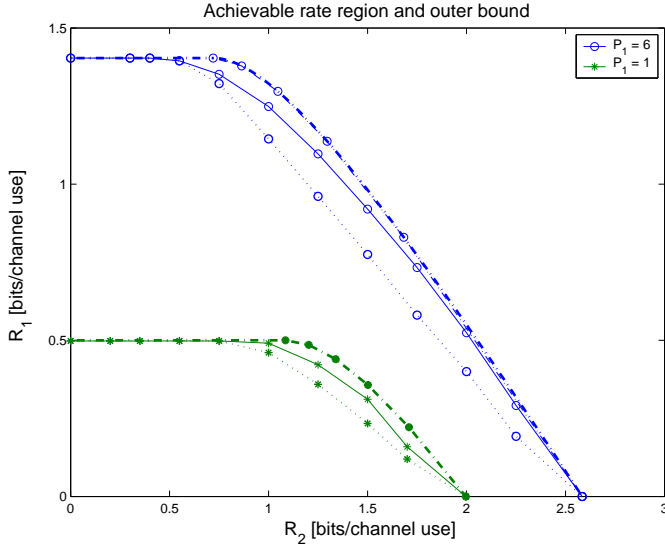


Fig. 7. Impact of reduced power of the cognitive transmitter to achievable rates. Rates achieved with Thm. 1 are shown in solid lines and rates of [6] are shown with dotted line. Dash-dotted line shows the outer bound.

When the encoder 2 does not transmit (i.e.  $P_2 = 0$ ), the channel reduces to the broadcast channel in which there is only the cooperating encoder communicating to the two receivers. The rates achieved in the BC are also shown. Unlike the BC channel rate region, the interference channel with one cooperating encoder region is flat for smaller values of  $R_2$ , reflecting the fact that for smaller values of  $R_2$  a cognitive transmitter does not need to cooperate. It can instead use its full power to precode and transmit  $W_1$  at the single-user rate as if the second user was not present. It starts cooperating only for higher  $R_2$ . At  $R_1 = 0$ , the cooperating encoder fully helps encoder 2, i.e.  $\alpha = 0$  and user 2 benefits from the coherent combining gain as indicated by the rate expression

$$R_{2,max} = \frac{1}{2} \log \left( 1 + \left( 1 + b \sqrt{\frac{P_1}{P_2}} \right)^2 P_2 \right).$$

The achievable rates come very close to the outer bound, especially for larger values of  $R_2$ , in the regime where the cognitive encoder dedicates more of its power to cooperate.

Fig. 6 shows achievable rates for different values of power  $P_2$  and fixed power  $P_1$ . As  $P_2$  decreases, the performance gets closer to the rate achieved in the BC with only the cooperating encoder transmitting to the two receivers. Since in the BC encoder 2 is not present, the rate region does not depend on  $P_2$  and is given by the dashed line. Fig. 7 shows the effect of reducing power at the cognitive encoder, keeping  $P_2$  constant. This has a higher impact, drastically reducing rate  $R_1$ .

For the Gaussian channel, the rates achieved with sequential encoding (48) can be evaluated for the choice of random variables  $X_2, X_1, U_{1a}, X_{1a}$  as in (49).  $U_{1c}$  carries a common message and is to be precoded against interference. Since the two channels from encoder 1 to the two receivers experience different interference, the carbon-copy method of [21] can be

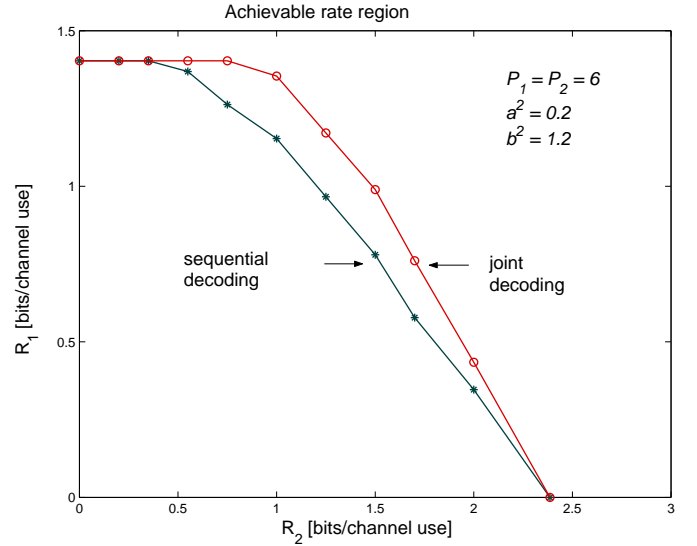


Fig. 8. Comparison of achievable rates with joint and sequential decoding.

used. More details on this approach are presented in [12].

Fig. 8 shows the performance of the two decoding schemes given the channel gain values  $a$  and  $b$  for which the performance differs significantly.

## VII. CONCLUSIONS AND FUTURE WORK

We have developed an encoding strategy for the interference channel with one cooperating encoder that generalizes previously proposed encoding strategies. We evaluated its performance and compared it to the performance of other schemes, focusing on the Gaussian channel. A comparison with [3] would be an interesting next step. It is unclear whether our strategy generalizes the scheme in [3], or whether a combination of the two techniques would achieve higher rates. We also compared the proposed scheme to the outer bound that we developed for the strong interference regime. We further developed a new outer bound that extends the Nair-El Gamal broadcast outer bound. Evaluating this bound for specific channels such as Gaussian may prove useful.

The cognitive radio channel shares some characteristics of both interference channels and broadcast channels. Combining encoding strategies developed for either of the two channel models therefore seems a natural approach. However, the optimality of a particular encoding scheme seems to be in part dictated by the channel conditions: for the Gaussian channel in which decoder 2 experiences weak interference, dirty-paper coding is capacity achieving. On the other hand, strong interference conditions may allow the cognitive receiver to decode the message not intended for him and therefore DPC against that message is not needed; superposition coding and rate-splitting achieve capacity. An even simpler scheme suffices when both receivers experience strong interference and can both decode the two messages. Neither DPC nor rate-splitting is needed; superposition coding is capacity-achieving. The encoding scheme presented in this paper is a combination

of rate-splitting, GP binning and superposition coding. We believe that this general encoding scheme may be capacity-achieving for certain special cases related to the channel or specific encoding/decoding constraints. Finding such special cases is a topic of ongoing investigation

#### APPENDIX A: PROOF OF THEOREM 1

*Proof:* (Theorem 1) **Code construction:** Ignore  $Q$ .

Choose a distribution  $p(x_{2a}, x_{2b}, u_{1c}, u_{1a}, x_1, x_2)$ .

- Split the rates as in (9)-(10).
- Generate  $2^{NR_{2a}}$  codewords  $x_{2a}^N(w_{2a})$ ,  $w_{2a} = 1, \dots, 2^{NR_{2a}}$ , by choosing  $x_{2a,n}(w_{2a})$  independently according to  $P_{X_{2a}}(\cdot)$ .
- For each  $w_{2a}$ : Generate  $2^{NR_{2b}}$  codewords  $x_{2b}^N(w_{2a}, w_{2b})$  using  $\prod_{n=1}^N P_{X_{2b}|X_{2a}}(\cdot|x_{2a,n}(w_{2a}))$ ,  $w_{2b} = 1, \dots, 2^{NR_{2b}}$ .
- For each pair  $(w_{2a}, w_{2b})$ : Generate  $x_2^N(w_{2a}, w_{2b})$  where  $x_2$  is a deterministic function of  $(x_{2a}, x_{2b})$ .
- Generate  $2^{N(R_{1c}+R_{1c'})}$  codewords  $u_{1c}^N(w_c, b_c)$ ,  $w_c = 1, \dots, 2^{NR_{1c}}$ ,  $b_c = 1, \dots, 2^{NR_{1c'}}$  using  $P_{U_{1c}}(\cdot)$ .
- For each  $u_{1c}^N(w_c, b_c)$ : Generate  $2^{N(R_{1a}+R_{1a'})}$  codewords  $u_{1a}^N(w_c, b_c, w_{1a}, b_{1a})$ ,  $w_{1a} = 1, \dots, 2^{NR_{1a}}$ ,  $b_{1a} = 1, \dots, 2^{NR_{1a'}}$  using  $\prod_{n=1}^N P_{U_{1a}|U_{1c}}(\cdot|u_{1c,n}(w_c, b_c))$ .
- For  $(w_1, w_2)$ : Generate  $x_1^N(w_{2a}, w_{2b}, w_c, b_c, w_{1a}, b_{1a})$  where  $x_1$  is a deterministic function of  $(x_{2a}, x_{2b}, u_{1c}, u_{1a}, x_2)$ .

Following the proof in [28, Appendix D], it can be shown that it is enough to choose respective  $x_2$  and  $x_1$  to be deterministic functions of  $(x_{2a}, x_{2b})$  and  $(x_{2a}, x_{2b}, u_{1c}, u_{1a}, x_2)$ .

**Encoders:** Encoder 1:

- 1) Split the  $NR_1$  bits  $w_1$  into  $NR_{1a}$  bits  $w_{1a}$  and  $NR_c$  bits  $w_c$ . Similarly, split the  $NR_2$  bits  $w_2$  into  $NR_{2a}$  bits  $w_{2a}$  and  $NR_{2b}$  bits  $w_{2b}$ . We write this as

$$w_1 = (w_{1a}, w_c), \quad w_2 = (w_{2a}, w_{2b}).$$

- 2) Try to find a bin index  $b_c$  so that  $(u_{1c}^N(w_c, b_c), x_{2a}^N(w_{2a}), x_{2b}^N(w_{2a}, w_{2b})) \in T_\epsilon(P_{U_{1c}X_{2a}X_{2b}})$  where  $T_\epsilon(P_{XY})$  denotes jointly  $\epsilon$ -typical set with respect to  $P_{XY}$ , see [29, Sect.8.6]. If no such  $b_c$  is found, choose  $b_c = 1$ .
- 3) For each  $(w_c, b_c)$ : Try to find a bin index  $b_{1a}$  such that  $(u_{1a}^N(w_c, b_c, w_{1a}, b_{1a}), u_{1c}^N(w_c, b_c), x_{2a}^N(w_{2a}), x_{2b}^N(w_{2a}, w_{2b})) \in T_\epsilon(P_{U_{1a}U_{1c}X_{2a}X_{2b}})$ . If cannot, choose  $b_{1a} = 1$ .
- 4) Transmit  $x_1^N(w_{2a}, w_{2b}, w_c, b_c, w_{1a}, b_{1a})$ .

Encoder 2: Transmit  $x_2^N(w_{2a}, w_{2b})$ .

**Decoders:** Decoder 1: Given  $y_1^N$ , choose  $(\hat{w}_c, \hat{b}_c, \hat{w}_{1a}, \hat{b}_{1a})$  if  $(u_{1c}^N(\hat{w}_c, \hat{b}_c), u_{1a}^N(\hat{w}_c, \hat{b}_c, \hat{w}_{1a}, \hat{b}_{1a}), y_1^N) \in T_\epsilon(P_{U_{1c}U_{1a}Y_1})$ . If there is more than one such a quadruple, choose one. If there is no such quadruple, choose  $(1, 1, 1, 1)$

Decoder 2: Given  $y_2^N$ , choose  $(\hat{w}'_{2a}, \hat{w}'_c, \hat{b}'_c, \hat{w}'_{2b})$  if  $(x_{2a}^N(\hat{w}'_{2a}), u_{1c}^N(\hat{w}'_c, \hat{b}'_c), x_{2b}^N(\hat{w}'_{2a}, \hat{w}'_{2b}), y_2^N) \in T_\epsilon(P_{X_{2a}U_{1c}X_{2b}Y_2})$ . If there is more than one such a quadruple, choose one. If there is no such quadruple, choose  $(1, 1, 1, 1)$ .

**Analysis:** Assume  $(w_{1a}, w_c, w_{2a}, w_{2b}) = (1, 1, 1, 1)$  was sent. Encoder error occurs if

- 1) Encoder 1 cannot find a bin index  $b_c$  such that  $(u_{1c}^N(1, b_c), x_{2a}^N(1), x_{2b}^N(1, 1)) \in T_\epsilon(P_{U_{1c}X_{2a}X_{2b}})$  which happens with probability

$$\begin{aligned} P_{e,enc1}^{(1)} &= P\left[\bigcap_{b_c=1}^{2^{NR_c}} (U_{1c}^N(1, b_c), x_{2a}^N(1), x_{2b}^N(1, 1)) \notin T_\epsilon(P_{U_{1c}X_{2a}X_{2b}})\right] \\ &= (1 - P[(U_{1c}^N(1, b_c), x_{2a}^N(1), x_{2b}^N(1, 1)) \in T_\epsilon(P_{U_{1c}X_{2a}X_{2b}})])^{2^{NR_c}} \\ &\leq (1 - (1 - \epsilon)2^{-N[I(U_{1c}; X_{2a}, X_{2b}) + \delta]})^{2^{NR_c}} \\ &\leq \exp(-(1 - \epsilon)2^{N[R'_c - I(U_{1c}; X_{2a}, X_{2b}) - \delta]}) \quad (52) \end{aligned}$$

where the first inequality follows from [29, Thm.8.6.1] and the second from  $(1-x)^m \leq \exp^{-mx}$ . From (52), the error probability  $P_{e,enc1}^{(1)}$  can be made arbitrarily small if

$$R'_c > I(U_{1c}; X_{2a}, X_{2b}) + \delta. \quad (53)$$

- 2) After it has determined bin index  $b_c$ , say  $b_c = 1$ , encoder 1 cannot find a bin index  $b_{1a}$  such that  $(u_{1a}^N(1, 1, 1, b_{1a}), u_{1c}^N(1, 1), x_{2a}^N(1), x_{2b}^N(1, 1)) \in T_\epsilon(P_{U_{1a}U_{1c}X_{2a}X_{2b}})$  which happens with probability

$$\begin{aligned} P_{e,enc1}^{(2)} &= P\left[\bigcap_{b_{1a}=1}^{2^{NR'_{1a}}} (U_{1a}^N(1, 1, 1, b_{1a}), u_{1c}^N(1, 1), x_{2a}^N(1), x_{2b}^N(1, 1)) \notin T_\epsilon(P_{U_{1a}U_{1c}X_{2a}X_{2b}})\right] \\ &= (1 - P[(U_{1a}^N(1, 1, 1, b_{1a}), u_{1c}^N(1, 1), x_{2a}^N(1), x_{2b}^N(1, 1)) \in T_\epsilon(P_{U_{1a}U_{1c}X_{2a}X_{2b}})])^{2^{NR'_{1a}}} \quad (54) \end{aligned}$$

We have

$$\begin{aligned} &P[(U_{1a}^N(1, 1, 1, b_{1a}), u_{1c}^N(1, 1), x_{2a}^N(1), x_{2b}^N(1, 1)) \in T_\epsilon(P_{U_{1a}U_{1c}X_{2a}X_{2b}})] \\ &= \sum_{u_{1a}^N \in T_\epsilon(P_{U_{1a}U_{1c}X_{2a}X_{2b}}|u_{1c}^N, x_{2a}^N, x_{2b}^N)} P[u_{1a}^N|u_{1c}^N] \\ &\geq (1 - \epsilon)2^{-N(H(U_{1a}|U_{1c}X_{2a}X_{2b}) - H(U_{1a}|U_{1c}) + \delta)} \\ &= (1 - \epsilon)2^{-N(I(U_{1a}; X_{2a}X_{2b}|U_{1c}) + \delta)} \quad (55) \end{aligned}$$

where the first inequality follows from the fact that  $U_{1a}^N$  was generated according to  $P_{U_{1a}|U_{1c}}$  (also directly from [30, Handout 3.]).

Employing (55), we can bound (54) as

$$\begin{aligned} P_{e,enc1}^{(2)} &\leq (1 - (1 - \epsilon)2^{-N[I(U_{1a}; X_{2a}, X_{2b}|U_{1c}) + \delta]})^{2^{NR'_{1a}}} \\ &\leq \exp(-(1 - \epsilon)2^{N[R'_{1a} - I(U_{1a}; X_{2a}, X_{2b}|U_{1c}) - \delta]}) \quad (56) \end{aligned}$$

We need

$$R'_{1a} > I(U_{1a}; X_{2a}, X_{2b}|U_{1c}) + \delta. \quad (57)$$



	Error event	Arbitrarily small positive error probability if
$E_1$	$(\hat{w}_c \neq 1, \hat{w}_{1a} = 1)$	$R_c + R'_c \leq I(U_{1c}, U_{1a}; Y_1)$
$E_2$	$(\hat{w}_c = 1, \hat{w}_{1a} \neq 1)$	$R_{1a} + R'_{1a} \leq I(U_{1a}; Y_1   U_{1c})$
$E_3$	$(\hat{w}_c \neq 1, \hat{w}_{1a} \neq 1)$	$R_c + R'_c + R_{1a} + R'_{1a} \leq I(U_{1c}, U_{1a}; Y_1)$
$E'_1$	$(\hat{w}'_{2a} \neq 1, \hat{w}'_{2b} = 1, \hat{w}'_c = 1)$	$R_{2a} \leq I(X_{2a}, X_{2b}; Y_2, U_{1c})$
$E'_2$	$(\hat{w}'_{2a} \neq 1, \hat{w}'_{2b} \neq 1, \hat{w}'_c = 1)$	$R_{2a} + R_{2b} \leq I(X_{2a}, X_{2b}; Y_2, U_{1c})$
$E'_3$	$(\hat{w}'_{2a} \neq 1, \hat{w}'_{2b} = 1, \hat{w}'_c \neq 1)$	$R_{2a} + R_c + R'_c \leq I(X_{2a}, X_{2b}, U_{1c}; Y_2) + I(U_{1c}; X_{2a}, X_{2b})$
$E'_4$	$(\hat{w}'_{2a} \neq 1, \hat{w}'_{2b} \neq 1, \hat{w}'_c \neq 1)$	$R_{2a} + R_{2b} + R_c + R'_c \leq I(X_{2a}, X_{2b}, U_{1c}; Y_2) + I(U_{1c}; X_{2a}, X_{2b})$
$E'_5$	$(\hat{w}'_{2a} = 1, \hat{w}'_{2b} \neq 1, \hat{w}'_c = 1)$	$R_{2b} \leq I(X_{2b}; Y_2, U_{1c}   X_{2a})$
$E'_6$	$(\hat{w}'_{2a} = 1, \hat{w}'_{2b} \neq 1, \hat{w}'_c \neq 1)$	$R_{2b} + R_c + R'_c \leq I(X_{2b}, U_{1c}; Y_2   X_{2a}) + I(U_{1c}; X_{2a}, X_{2b})$

TABLE I  
ERROR EVENTS IN JOINT DECODING AND CORRESPONDING RATE BOUNDS.

**Decoder errors:** Possible error events at decoders are shown in the first column of Table I. We next derive the corresponding rate bounds given in the second column of the same table, which guarantee that the error probability of each event can be made small as  $N$  gets large. Bounds for  $E_1, E'_1, E'_3$  are loose. The rest of the rate expressions in Table I yield (11)-(16).

Consider the probability of event  $E_1$  :

$$P[\hat{w}_c \neq 1, \hat{w}_{1a} = 1] = \sum_{w_c=2}^{2^{NR_c}} \sum_{b_c=1}^{2^{NR'_c}} P[(U_{1c}^N(w_c, b_c), U_{1a}^N(w_c, b_c, 1, 1), Y_1^N) \in T_\epsilon(P_{U_{1c}U_{1a}Y_1})] \leq 2^{-N[I(U_{1c}, U_{1a}; Y_1) - (R_c + R'_c) - \delta]} \quad (58)$$

by [29, Thm.8.6.1] and [30, Handout 1, Thm. 2]. From (58), the arbitrarily small error probability of  $E_1$  requires

$$R_c + R'_c < I(U_{1c}, U_{1a}; Y_1). \quad (59)$$

Similarly, the probability of  $E_2$  is

$$P[\hat{w}_c = 1, \hat{w}_{1a} \neq 1] = \sum_{w_{1a}=2}^{2^{NR_{1a}}} \sum_{b_{1a}=1}^{2^{NR'_{1a}}} P[(U_{1c}^N(1, 1), U_{1a}^N(1, 1, w_{1a}, b_{1a}), Y_1^N) \in T_\epsilon(P_{U_{1c}U_{1a}Y_1})] \leq 2^{-N[I(U_{1a}; Y_1 | U_{1c}) - (R_{1a} + R'_{1a}) - \delta]} \quad (60)$$

where the inequality follows by [30, Thm., Handout 3]. We need

$$R_{1a} + R'_{1a} < I(U_{1a}; Y_1 | U_{1c}). \quad (61)$$

The probability of  $E_3$  is, similarly as in (58),

$$P[\hat{w}_c \neq 1, \hat{w}_{1a} \neq 1] = \sum_{w_c=2}^{2^{NR_c}} \sum_{b_c=1}^{2^{NR'_c}} \sum_{w_{1a}=2}^{2^{NR_{1a}}} \sum_{b_{1a}=1}^{2^{NR'_{1a}}} P[(U_{1c}^N(w_c, b_c), U_{1a}^N(w_c, b_c, w_{1a}, b_{1a}), Y_1^N) \in T_\epsilon(P_{U_{1c}U_{1a}Y_1})] \leq 2^{-N[I(U_{1c}, U_{1a}; Y_1) - (R_c + R'_c + R_{1a} + R'_{1a}) - \delta]} \quad (62)$$

requiring

$$R_c + R'_c + R_{1a} + R'_{1a} < I(U_{1c}, U_{1a}; Y_1). \quad (63)$$

We next consider the error events at decoder 2. For  $E'_1$

$$P[\hat{w}'_{2a} \neq 1, \hat{w}'_{2b} = 1, \hat{w}'_c = 1] = \sum_{w_{2a}=2}^{2^{NR_{2a}}} P[(U_{1c}^N(1, 1), X_{2a}^N(w_{2a}), X_{2b}^N(w_{2a}, 1), Y_2^N) \in T_\epsilon(P_{U_{1c}X_{2a}X_{2b}Y_2})]. \quad (64)$$

We have

$$P[(U_{1c}^N(1, 1), X_{2a}^N(w_{2a}), X_{2b}^N(w_{2a}, 1), Y_2^N) \in T_\epsilon(P_{U_{1c}X_{2a}X_{2b}Y_2})] = \sum_{(u_{1c}^N, x_{2a}^N, x_{2b}^N, y_2^N) \in T_\epsilon} P[x_{2a}^N, x_{2b}^N] P[y_2^N | u_{1c}^N] \leq 2^{-N[I(X_{2a}, X_{2b}; Y_2, U_{1c}) - \delta]}. \quad (65)$$

From (64) and (65),

$$R_{2a} < I(X_{2a}, X_{2b}; Y_2, U_{1c}). \quad (66)$$

The probability of event  $E'_2$  is

$$P[\hat{w}'_{2a} \neq 1, \hat{w}'_{2b} \neq 1, \hat{w}'_c = 1] = \sum_{w_{2a}=2}^{2^{NR_{2a}}} \sum_{w_{2b}=2}^{2^{NR_{2b}}} P[(U_{1c}^N(1, 1), X_{2a}^N(w_{2a}), X_{2b}^N(w_{2a}, w_{2b}), Y_2^N) \in T_\epsilon(P_{U_{1c}X_{2a}X_{2b}Y_2})]. \quad (67)$$

Following the same steps as in (65) and using (67) it can be shown that the arbitrarily small error probability of  $E'_2$  requires

$$R_{2a} + R_{2b} < I(X_{2a}, X_{2b}; Y_2, U_{1c}). \quad (68)$$

We next consider  $E'_3$ :

$$P[\hat{w}'_{2a} \neq 1, \hat{w}'_{2b} = 1, \hat{w}'_c \neq 1] = \sum_{w_{2a}=2}^{2^{NR_{2a}}} \sum_{w_c=2}^{2^{NR_c}} \sum_{b_c=1}^{2^{NR'_c}} P[(U_{1c}^N(w_c, b_c), X_{2a}^N(w_{2a}), X_{2b}^N(w_{2a}, 1), Y_2^N) \in T_\epsilon(P_{U_{1c}X_{2a}X_{2b}Y_2})]. \quad (69)$$

$$X_{2b}^N(w_{2a}, 1), Y_2^N) \in T_\epsilon(P_{U_{1c}X_{2a}X_{2b}Y_2})]. \quad (70)$$

We have

$$P[(U_{1c}^N(w_c, b_c), X_{2a}^N(w_{2a}), X_{2b}^N(w_{2a}, 1), Y_2^N) \in T_\epsilon(P_{U_{1c}X_{2a}X_{2b}Y_2})] = \sum_{(u_{1c}^N, x_{2a}^N, x_{2b}^N, y_2^N) \in T_\epsilon} P[x_{2a}^N, x_{2b}^N] P[u_{1c}^N] P[y_2^N] \leq 2^{-N[I(X_{2a}, X_{2b}, U_{1c}; Y_2) + I(U_{1c}; X_{2a}, X_{2b}) - \delta]}. \quad (71)$$

From (70) and (71) it follows that

$$R_{2a} + R_c + R'_c < I(X_{2a}, X_{2b}, U_{1c}; Y_2) + I(U_{1c}; X_{2a}, X_{2b}). \quad (72)$$

For  $E'_4$  we use the same approach as in (70) and reuse (71) to obtain

$$R_{2a} + R_{2b} + R_c + R'_c < I(X_{2a}, X_{2b}, U_{1c}; Y_2) + I(U_{1c}; X_{2a}, X_{2b}). \quad (73)$$

We continue by considering error event  $E'_5$ :

$$P[\hat{w}'_{2a} = 1, \hat{w}'_{2b} \neq 1, \hat{w}'_c = 1] = \sum_{w_{2b}=2}^{2^{NR_{2b}}} P[(U_{1c}^N(1, 1), X_{2a}^N(1), X_{2b}^N(1, w_{2b}), Y_2^N) \in T_\epsilon(P_{U_{1c}X_{2a}X_{2b}Y_2})]. \quad (74)$$

We have

$$\begin{aligned} P[(U_{1c}^N(1, 1), X_{2a}^N(1), X_{2b}^N(1, w_{2b}), Y_2^N) \in T_\epsilon(P_{U_{1c}X_{2a}X_{2b}Y_2})] \\ = \sum_{(u_{1c}^N, x_{2a}^N, x_{2b}^N, y_2^N) \in T_\epsilon} P[x_{2a}^N, x_{2b}^N] P[u_{1c}^N, y_2^N | x_{2a}^N] \\ \leq 2^{-N[I(X_{2b}, U_{1c}; Y_2 | X_{2a}) - \delta]}. \end{aligned} \quad (75)$$

From (74) and (75) it follows that

$$R_{2b} < I(X_{2b}, U_{1c}; Y_2 | X_{2a}). \quad (76)$$

For the error event  $E'_6$  we have

$$\begin{aligned} P[\hat{w}'_{2a} = 1, \hat{w}'_{2b} \neq 1, \hat{w}'_c \neq 1] = \sum_{w_{2b}=2}^{2^{NR_{2b}}} \sum_{w_c=2}^{2^{NR_c}} \sum_{b_c=1}^{2^{NR'_c}} P[(U_{1c}^N(w_c, b_c), X_{2a}^N(1), X_{2b}^N(1, w_{2b}), Y_2^N) \in T_\epsilon(P_{U_{1c}X_{2a}X_{2b}Y_2})]. \end{aligned} \quad (77)$$

Again

$$\begin{aligned} P[(U_{1c}^N(w_c, b_c), X_{2a}^N(1), X_{2b}^N(1, w_{2b}), Y_2^N) \in T_\epsilon(P_{U_{1c}X_{2a}X_{2b}Y_2})] \\ = \sum_{(u_{1c}^N, x_{2a}^N, x_{2b}^N, y_2^N) \in T_\epsilon} P[x_{2a}^N, x_{2b}^N] P[u_{1c}^N] P[y_2^N | x_{2a}^N] \\ \leq 2^{-N[I(X_{2b}, U_{1c}; Y_2 | X_{2a}) + I(U_{1c}; X_{2a}, X_{2b}) - \delta]}. \end{aligned} \quad (78)$$

From (77) and (78) it follows that

$$R_{2b} + R_c + R'_c < I(X_{2b}, U_{1c}; Y_2 | X_{2a}) + I(U_{1c}; X_{2a}, X_{2b}). \quad (79)$$

## APPENDIX B: PROOF OF LEMMA 1

*Proof:* (Lemma 1) For  $I(S; Y, U) \leq R_s \leq H(S)$  use G-P coding [13]. The achieved rate is

$$R \leq I(U; Y) - I(U; S). \quad (80)$$

Note that  $I(U; Y) - I(U; S) \leq I(X; Y | S)$ .

For  $R_s < I(S; Y, U)$  proceed as follows.

**Code construction:** For every codeword  $s^N(j)$ ,  $j = 1, \dots, 2^{NR_s}$  generate  $2^{NR}$  codewords  $u^N(w, j)$ ,  $w = 1, \dots, 2^{NR}$  using  $\prod_{n=1}^N P_{U|S}(\cdot | s_n(j))$ .

**Encoder:** Given  $w$  and  $s^N(j)$ , choose  $u^N(w, j)$  and transmit  $x^N = f^N(u^N(w, j), s^N(j))$ .

**Decoder:** Given  $y^N$ , try to find  $(w, j)$  such that  $(u^N(w, j), s^N(j), y^N) \in T_\epsilon(P_{SUY})$ .

**Analysis:** Suppose  $w = 1, j = 1$  was sent. Error  $\{\hat{w} \neq 1\}$  occurs if  $\{\hat{w} \neq 1, \hat{j} \neq 1\}$  or  $\{\hat{w} \neq 1, \hat{j} = 1\}$ . The probability of error is

$$\begin{aligned} P_e &= \sum_{j=2}^{2^{NR_s}} \sum_{w=1}^{2^{NR}} P[(U^N(w, j), S^N(j), Y^N) \in T_\epsilon(P_{SUY})] \\ &\quad + \sum_{w=2}^{2^{NR}} P[(U^N(w, 1), s^N(1), Y^N) \in T_\epsilon(P_{SUY})] \\ &\leq 2^{N(R+R_s-I(US;Y)+\delta_1)} + 2^{N(R-I(U;Y|S)+\delta_2)} \end{aligned} \quad (81)$$

where  $\delta_1, \delta_2 \rightarrow 0$  as  $N \rightarrow \infty$ . From (81) and  $I(U; Y | S) = I(X; Y | S)$  it follows that

$$R \leq \min\{I(U, S; Y) - R_s, I(X; Y | S)\}. \quad (82)$$

Note that we could have chosen  $u^N = x^N$  in the superposition coding above, so that (82) is

$$R \leq \min\{I(X, S; Y) - R_s, I(X; Y | S)\}. \quad (83)$$

■

## APPENDIX C: PROOF OF THEOREM 4

*Proof:* (Theorem 4) Consider a code  $(M_1, M_2, N, P_e)$  for the interference channel with one cooperating encoder. We first consider the bound (31). Fano's inequality implies that for reliable communication we require

$$\begin{aligned} N(R_1 + R_2) &\leq I(W_1; Y_1^N) + I(W_2; Y_2^N) \\ &\leq^{(a)} I(W_1; Y_1^N | W_2) + I(W_2; Y_2^N) \\ &= \sum_{i=1}^N I(W_1; Y_1^i | W_2, Y_{2,i+1}^N) - I(W_1; Y_1^{i-1} | W_2, Y_{2,i}^N) \\ &\quad + I(W_2; Y_{2,i} | Y_{2,i+1}^N) \\ &= \sum_{i=1}^N I(W_1; Y_1^i | W_2, Y_{2,i+1}^N) - [I(W_1, Y_{2,i}; Y_1^{i-1} | W_2, Y_{2,i+1}^N) \\ &\quad - I(Y_{2,i}; Y_1^{i-1} | W_2, Y_{2,i+1}^N)] + I(W_2; Y_{2,i} | Y_{2,i+1}^N) \\ &=^{(b)} \sum_{i=1}^N I(W_1; Y_{1,i} | W_2, V_i) - I(Y_{2,i}; Y_1^{i-1} | W_1, W_2, Y_{2,i+1}^N) \\ &\quad + I(W_2, Y_1^{i-1}; Y_{2,i} | Y_{2,i+1}^N) \\ &\leq \sum_{i=1}^N I(W_1; Y_{1,i} | W_2, V_i) + I(W_2, V_i; Y_{2,i}) \end{aligned} \quad (84)$$

where (a) follows from the independence of  $W_1, W_2$ ; in (b), we let  $Y_{t,i}^j = (Y_{t,i}, \dots, Y_{t,j})$  and  $V_i = [Y_1^{i-1}, Y_{2,i+1}^N]$ .

We next consider the bound (29). Fano's inequality implies

$$\begin{aligned}
NR_2 &\leq I(W_2; Y_2^N) \\
&= \sum_{i=1}^N I(W_2; Y_{2,i} | Y_{2,i+1}^N) \\
&\leq \sum_{i=1}^N I(W_2, Y_1^{i-1}, Y_{2,i+1}^N; Y_{2,i}) \\
&= \sum_{i=1}^N I(W_2, V_i; Y_{2,i}). \tag{85}
\end{aligned}$$

Note that for (84)-(85) we have used only the independence of  $W_1$  and  $W_2$ , and the non-negativity of mutual information. The bounds (28) and (30) thus follow by symmetry.

We introduce random variables  $U_{1,i} = W_1$  and  $U_{2,i} = W_2$  for all  $i$ , to get the bounds in the form (28)-(31). Observe that  $U_{1,i}$  and  $U_{2,i}$  are independent. Furthermore, due to unidirectional cooperation, the joint probability distribution factors as in (32). ■

#### APPENDIX D: PROOF OF THEOREM 5

*Proof:* (Theorem 5) The bound (42) follows by standard methods. To prove (43), consider (31) and

$$\begin{aligned}
I(U_1; Y_1 | U_2, V) &\leq I(U_1; Y_1, X_2 | U_2, V) \\
&= I(U_1; Y_1 | U_2, V, X_2) \\
&\leq I(U_1, X_1; Y_1 | U_2, V, X_2) \\
&= I(X_1; Y_1 | U_2, V, X_2) \\
&\leq I(X_1; Y_2 | U_2, V, X_2) \tag{86}
\end{aligned}$$

where the second step follows by the Markov chain (32), and the last step follows by (41). We similarly have

$$I(V, U_2; Y_2) \leq I(U_2, V, X_2; Y_2). \tag{87}$$

Combining inequalities (31), (86) and (87) gives (43). ■

#### REFERENCES

- [1] H. Sato, "Two user communication channels," *IEEE Trans. Inf. Theory*, vol. 23, no. 3, pp. 295–304, May 1977.
- [2] A. B. Carleial, "Interference channels," *IEEE Trans. Inf. Theory*, vol. 24, no. 1, pp. 60–70, Jan. 1978.
- [3] N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive radio channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 1813–1827, May 2006.
- [4] J. Mitola, *Cognitive Radio Architecture*. John Wiley Sons, Inc., 1991.
- [5] N. Devroye, P. Mitran, and V. Tarokh, "Limits on communications in a cognitive radio channel," in *IEEE Comm. Magazine*, vol. 44, no. 6, June 2006, pp. 44–49.
- [6] J. Jiang and Y. Xin, "On the achievable rate regions for interference channels with degraded message sets," *IEEE Trans. Inf. Theory*, submitted, Apr. 2007.
- [7] W. Wu, S. Vishwanath, and A. Arapostathis, "On the capacity of Gaussian weak interference channels with degraded message sets," in *Proc. Conf. Inf. Sciences and Systems (CISS)*, also submitted to *IEEE Trans. Inf. Theory*, Mar. 2006.
- [8] A. Jovičić and P. Viswanath, "Cognitive radio: An information-theoretic perspective," in *Proc. IEEE Int. Symp. Inf. Theory*, also submitted to *IEEE Trans. Inf. Theory*, <http://www.arxiv.org/pdf/0604/0604107v2.pdf>, July 2006, pp. 2413–2417.
- [9] S. Sridharan and S. Vishwanath, "On the capacity of a class of MIMO cognitive radios," in *IEEE Information Theory Workshop (ITW 2007)*, Lake Tahoe, Ca, USA, Sept. 2007.
- [10] D. Tuninetti, "The interference channels with generalized feedback," in *IEEE Proc. Int. Symp. Inf. Th.*, jun 2007.
- [11] Y. Cao, B. Chen, and J. Zhang, "A new achievable rate region for interference channels with common information," in *Proc. IEEE Wireless Comm. and Networking Conf.*, 2007.
- [12] I. Marić, A. Goldsmith, G. Kramer, and S. Shamai(Shitz), "On the capacity of interference channels with a cognitive transmitter," in *Information Theory and Applications (ITA)*, <http://ita.ucsd.edu/workshop/07/files/paper/paper431.pdf>, Jan. 2007.
- [13] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Problemy Peredachi Informatsii*, vol. 9, no. 1, pp. 19–31, 1980.
- [14] C. Nair and A. E. Gamal, "An outer bound to the capacity region of the broadcast channel," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 350–355, Jan. 2007.
- [15] I. Marić, R. D. Yates, and G. Kramer, "Capacity of interference channels with partial transmitter cooperation," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, Oct. 2007.
- [16] T. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 49–60, Jan. 1981.
- [17] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.
- [18] H. Weingarten, Y. Steinberg, and S. Shamai(Shitz), "The capacity region of the Gaussian multiple input multiple output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936–3964, Sept. 2006.
- [19] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates and sum rate capacity of gaussian MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.
- [20] P. Mitran, N. Devroye, and V. Tarokh, "On compound channels with side-information at the transmitter," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1745–1755, Apr. 2006.
- [21] A. Khisti, U. Erez, A. Lapidoth, and G. W. Wornell, "Carbon copying onto dirty paper," *IEEE Trans. Inf. Theory*, vol. 53, no. 5, pp. 1814–1827, May 2007.
- [22] J. Jiang, Y. Xin, and H. Garg, "Interference channels with common information," *IEEE Trans. Inf. Theory*, submitted.
- [23] Y. Liang, A. Somekh-Baruch, V. Poor, S. Shamai(Shitz), and S. Verdú, "Cognitive interference channels with confidential messages," in *45th Annual Allerton Conference on Communication, Control and Computing*, Allerton House, Monticello, IL, USA, Sept. 2007.
- [24] T. Cover, Y. Kim, and A. Suvivong, "Channel capacity and state uncertainty reduction for state-dependent channels," in *Proc. IEEE Int. Symp. Inf. Theory*, June 2007.
- [25] —, "Simultaneous communication of data and state," <http://arxiv.org/PScache/cs/pdf/0703/0703005v1.pdf>, 2007.
- [26] A. Jovičić and P. Viswanath, "Cognitive radio: An information-theoretic perspective," *IEEE Trans. Inf. Theory*, submitted, <http://www.arxiv.org/pdf/cs.IT/0604107.pdf>, 2006.
- [27] S. Lall, *Advanced Topics in Computation for Control*. Lecture notes, Stanford University, 2004.
- [28] F. M. J. Willems, "The discrete memoryless multiple access channel with cribbing encoders," *IEEE Trans. Inf. Theory*, vol. 31, no. 3, pp. 313–327, May 1985.
- [29] T. Cover and J. Thomas, *Elements of Information Theory*. John Wiley Sons, Inc., 1991.
- [30] G. Kramer, *Multiple User Information Theory*, lecture notes. unpublished, 2006.